EXPORTIT MARKET BRIEF: PEER-TO-PEER TECHNOLOGIES

Raymond Cho
February, 2002
U.S. Department of Commerce
Office of Information Technologies
http://exportit.ita.doc.gov

"There are a lot of neat things coming out of this peer-to-peer model."
- Bill Gates, Chairman of Microsoft Corporation

"Peer-to-peer computing could be as important to the Internet's future as the Web browser was to its past."
- Pat Gelsinger, VP and Chief Technology Officer of the Intel Architecture Group

Sub-Sector Definition

Peer-to-peer computing ("P2P") is the sharing of computer resources by direct exchange between systems. Types of resources include information, processing cycles, cache storage, and disk storage for files. While in the client-server model certain systems are dedicated to serving other systems, in the decentralized P2P model each system or peer has equivalent capabilities and responsibilities. The P2P architecture essentially transforms client systems from simple service recipients to service providers as well.

There are two main types of P2P networks:

File-sharing

This is the direct exchange of files between two or more peers on a P2P network. Files are not retrieved nor stored on a central server to which client systems connect, but rather are transferred directly from the system on which they reside.

There are two types of file-sharing networks: centralized and decentralized. The best example of the centralized type is the Napster service. Napster's central servers act as matchmakers – taking a file request from a user, checking its internal lists for another user that possesses that file, and connecting the requesting user with the second for a direct file transfer. In the centralized file-sharing network, servers mediate and facilitate peer-to-peer transfers.

The Gnutella search protocol best represents the decentralized file-sharing model. A user's file request is first sent to several other user systems. If the file is present on any of these systems, then a direct transfer is established; if not, then each of those systems forwards the request to several other systems, and so on until the file is found. File searches can quickly become exponential: a request goes to four systems, then to sixteen, then to sixty-four, etc. The decentralized model bypasses intermediary servers and transforms all systems into both servers and clients.

Distributed computing

In this model, all the servers, desktops, notebooks, etc. that constitute a network become peers that contribute all or part of their resources, such as processing cycles or storage, to an overall computing effort. Unused or wasted resources can be harnessed to collectively surpass the normal available power of a single system. For example, large computational jobs can be distributed across thousands of idle Internet-connected personal computers to

create the equivalent of a virtual supercomputer. The distributed computing architecture has become possible due to increasingly inexpensive computing power, improved bandwidth, and increased storage capacities. SETI@Home, the project to analyze signals from a radio telescope in Puerto Rico for signs of extraterrestrial life, has over 2 million computers worldwide contributing processing power and is perhaps the best-known example of the distributed computing P2P model.

Industry Players and Technology and Product Developments

The P2P space is rapidly evolving and contains a core group of first-movers as well as a growing number of established larger companies. Both the file-sharing and distributed computing models, as well as variant approaches, are represented by an array of primarily American companies. Many of these companies are pushing the boundaries of both technology and the law, and few if any have achieved profitability. File-sharing applications like Napster, for instance, have raised copyright protection and intellectual property rights issues. Various distributed computing efforts, such as the Intel-United Devices Cancer Research Project, are tapping into a large global pool of unused processing cycles to cost-effectively undertake charitable or scientific research. Some companies, like Groove Networks or Applied MetaComputing, are attempting to apply P2P concepts to the enterprise.

Due in large measure to its popularity and the litigation surrounding it, Napster is probably the best known of all P2P companies. The Napster software application -- created by a freshman at Northeastern University – enables the peer-to-peer transmission of music files across the Internet, often without payment to copyright holders. Recent court decisions have effectively stopped the ability of users to freely share copyrighted materials using Napster and the company is attempting to implement payment-based systems. At the time of this writing, the Napster service was not fully operational.

Like Napster, Gnutella also enables P2P file sharing. However, Gnutella is not a company but rather a protocol that can be integrated into software applications to make peer-to-peer communications possible. Since Gnutella is designed to be free of any central authority and the name doesn't refer to a company or even a specific application, it has thus far avoided the legal scrutiny of the Napster service. Many services and/or software programs that incorporate the Gnutella technology have emerged, such as Windows-based BearShare, LimeWire, and ToadNode. These applications are sometimes called "servents," a word conjoining both "server" and "servant" in a pun that plays off the P2P concept.

In addition to Napster and Gnutella, other file-sharing systems and software in the P2P space include Aimster, Freenet, and MojoNation. Aimster uses America Online's (AOL) instant messaging service to allow users to send files to one another. Freenet, MojoNation, and the now defunct Scour were Napster-like services that facilitated the transfer of all types of files, not just for sound and music. The P2P area is rapidly evolving and as it does, companies are both emerging and falling by the wayside. An increasing number of larger companies are coming out with P2P-related products and services, a development that may be heralding a maturation of the space.

A few of the early entrants into the space include IBM, Intel, and Hewlett-Packard, which together in late 2000 launched the Peer-to-Peer Working Group, which aims to standardize and commercialize P2P technologies. Sun, Microsoft, and others have also been supportive of P2P. Sun is working on a standard development platform for peer-to-peer applications called JXTA (pronounced "Juxta"), while Microsoft has expressed interest in the technology and has incorporated some P2P-like concepts into its .Net initiative but has not yet released any purely P2P-based products or services. In practice, P2P technology generally favors a "fat client" (one in which most of the processing power is on the desktop) versus a "thin client" (where most processing is done on a central server) computing approach. For the most part, enthusiasm for P2P has correlated well with a firm's stake in one of these two approaches.

Intel, which is the largest supplier of processor chips for fat clients (PCs), has been especially active on the P2P issue. The company claims to have saved over $500 million internally by capturing spare computing cycles through a P2P application it calls "Netbatch," and is also planning to implement a P2P-based multimedia training system for its employees worldwide. Intel has made significant investments in several P2P start-ups, such as Uprizer, Groove Networks, and others. Meanwhile, a company that has in the past been viewed as a major proponent of thin client environments, Sun Microsystems, is positioning its JXTA initiative as the standard development platform for P2P applications. JXTA incorporates elements that Sun believes are critical to P2P networking, such as communication protocols, dynamic group directories, management, and security. Like Intel, Sun has also been acquiring or investing in smaller companies with innovative technologies; in early 2001 they acquired P2P start-up InfraSearch, whose distributed search technology will be integrated into the JXTA project.

While a lot of the corporate activity surrounding P2P has focused on the file-sharing aspects of the technology, there have also been some major initiatives to take advantage of the distributed computing power of P2P. IBM has staked out space in that area through its "grid computing" project, which envisions a world in which anyone at a desktop or hand-held computer could have the power of a supercomputer available to them via a distributed computing architecture. Grid computing also includes the ability to tap into large database files and application programs across high-speed network connections. IBM has already won several major contracts to implement a grid computing solution, including national projects in the United Kingdom and in the Netherlands. Microsoft, Sun, and others are also working on grid computing initiatives. Microsoft is supporting research on putting grid computing tools in Microsoft's software like Windows and it's .Net web services software. Sun is putting forward its grid computing software as an open-source project, which makes the code freely available for programmers around the world to work on.

Although major U.S. IT companies are still staking out territory in the P2P space, it appears likely that many will have divergent visions of the best way to approach the technology. Sun's JXTA platform may come into conflict with Intel and IBM's Peer-to-Peer Working Group, which itself was created to outline P2P standards. Some elements

of JXTA will also become part of Sun's Open Net Environment, a software suite for building Web-based applications that is being positioned against Microsoft's software initiative .Net. Many of the efforts underway involve carving out new markets where none exist currently. Niches are being filled by smaller companies with new approaches to old dilemmas. For example, online B2B marketplaces are being established that use peer-to-peer technology to connect buyers and sellers and bypass centralized marketplace operators. As P2P continues to develop, it still remains to be seen what approaches will emerge as viable business models.

Potential Technology Issues

Security issues are becoming a major concern as the P2P space matures. Early file-sharing applications like Napster showed the relative ease with which copyrighted intellectual property could be moved around networks, and companies are wary of implementing broad P2P file-sharing solutions without ensuring that their data does not get into the wrong hands. While many IT vendors are working on P2P security and authentication solutions, most still recommend that companies encrypt their data when using P2P networks. More robust P2P approaches, such as distributed or grid computing, also raise security concerns. Access to databases and processors across the grid, while potentially a powerful tool, also creates an opportunity for hacker or virus intrusions.

The network infrastructure could be another major hurdle for P2P. The widespread adoption of P2P would mean a large volume of data in transit, which would require a great deal more bandwidth; some estimates say that up to a thousand times more network capacity will be needed. Napster and other early file-sharing applications put this issue into perspective in 2000 when many universities across the country banned its use for putting their networks under incredible strain. For example, at Indiana University, the use of Napster was consuming 50 percent of the total university network bandwidth. Some innovative interim remedies have been proposed for possible network congestion that may prove helpful before a large network build-out is complete. For example, a "super-peer" solution that has emerged for the Gnutella network which would connect groups of slower computers to faster systems. Searches rippling through the network would be filtered only through the fast super-peer system and would skip slow machines. Other solutions providers are working on software that can "add intelligence" to the way Gnutella and other P2P systems route traffic and use available bandwidth.

Another important infrastructure component is the software, which is both the most important enabling technology for grid computing and its biggest problem. In the grid computing space, software must manage and coordinate the sharing of databases, applications, and computing power across the network, and do so reliably and securely. What is emerging as one of the key standards for grid computing software comes from the Globus project, a development effort led by the Argonne National Laboratory in France and the University of Southern California's Information Sciences Institute. The Globus project, started in 1996, is developing its software following the open-source model, in which computer code is openly shared, allowing programmers to modify, improve, and fix the software. It is the same approach to software development used by the Apache Web server project and Linux.

Reliability and other management issues will also play role in how successful P2P will be in the enterprise. For instance, someone could turn his or her computer off in the middle of a process when others are attempting to access the information. The increasing use of laptops means that vital corporate data could be disconnected from the network at any time. Regular data backups are difficult when individual machines can be turned off. Currently these problems are addressed more effectively by centralized servers, which are constantly connected and more reliable.

Market Size, Demand Factors, and Market Forecasts

Peer-to-peer is an emerging and diverse area and there are currently few concrete numbers available on sales or usage. Market research firm Frost and Sullivan projects that corporate users of P2P technologies will rise from 60,000 currently to 6.2 million by 2007, and that the enterprise P2P market will increase from $42.8 million today to $4.53 billion in 2007. The Gartner Group estimates that by 2003, over 30 percent of all U.S. corporations will be experimenting with peer-to-peer. One estimate of the potential for subscriptions in the content file-sharing space, represented by Gnutella and others, places the market size at over $1.5 billion by 2005. Currently, however, the free exchange of often-copyrighted files over P2P networks such as Gnutella continues apace. WebNoize estimates that over 3 billion files were transferred over the leading P2P file-sharing systems in August 2001, more than were downloaded via Napster at its February 2001 peak of 2.79 billion.

As the technology continues to mature, an increasing number of large companies are beginning to adopt P2P-based solutions to conduct business more cheaply and effectively. Firms like Intel, GlaxoSmithKline, Raytheon, Ernst & Young, and First Union Bank are turning towards P2P to help them manage and share information across distances or as a collaborative tool. For example, First Union is using a distributed computing system from New York start-up DataSynapse to harness unused processing power from existing computer systems and to improve their existing distribution architecture. Raytheon, GlaxoSmithKline, and the U.S. Department of Defense are implementing file-sharing software from Groove Networks, a P2P company headed by Lotus Notes creator Ray Ozzie. Governments in other countries, such as the United Kingdom and the Netherlands, are using P2P systems to facilitate and enhance scientific research efforts.

There are still potential applications of P2P yet to be explored. For example, the U.S. Army is developing ways to integrate wireless peer-to-peer technology into its training methods, and is also considering using P2P networking in real military combat situations. The focus of the army's research is on ways of improving battle simulations. It believes that by linking together hundreds of soldiers, each equipped with a head-mounted display that broadcasts details of a virtual environment, it would be possible for military units to accurately simulate various scenarios on the fly. A squad could practice taking a hill virtually minutes before they actually conducted the operation. New and innovative ways to use P2P such as this will make it harder to assess the entire market landscape.

Whether P2P technologies gain widespread traction in the enterprise is contingent on the success of these early adopters. Like client-server before it, P2P must prove itself as worthy of the additional expense and integration problems. Moreover, some technologies, like grid computing, are still in the early developmental stages and will require a more robust network or software infrastructure supporting them before they can be fully and economically adopted by many enterprises. Despite these challenges, many observers see a bright future for P2P, particularly as a powerful tool in workgroup collaboration systems such as personal-area networks in which workers can connect PCs on the fly to create workgroups in environments where there is no formal network. Other areas where P2P is likely to make an impact are supply-chain management and resource sharing.